

Μέθοδοι Πυρήνα (Kernel Methods)

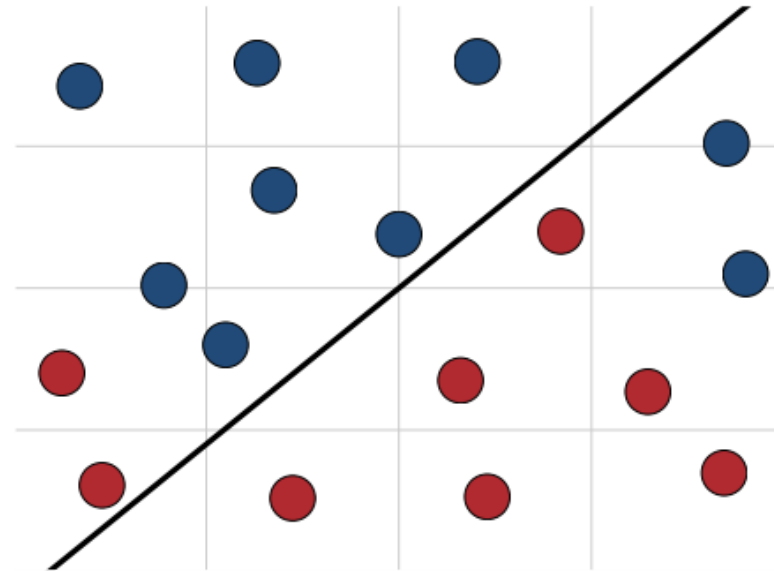
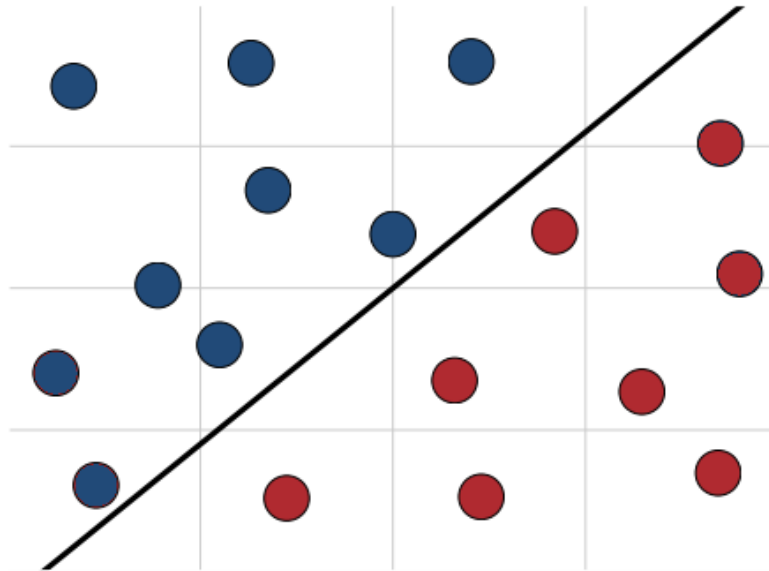
Εισαγωγή

Μοντέλα Ταξινόμησης

1. **Παραγωγικά** (*Generative*)
 - Προσπαθούν να ανακαλύψουν τις στατιστικές ιδιότητες των δεδομένων (*υποκείμενη κατανομή*) που ανήκουν σε μια κλάση
 - Παραδείγματα
 - Γκαουσιανά μοντέλα μείξης, αφελής μπεϋζιανός ταξινομητής, ...
2. **Διαχωριστικά** (*Discriminative*)
 - Προσπαθούν να εντοπίσουν *υπερ-επίπεδα διαχωρισμού* που *τεμαχίζουν* το χώρο των δεδομένων σε *περιοχές* εντός των οποίων βρίσκονται δείγματα *μιας μόνο* κλάσης
 - Παραδείγματα
 - Λογιστική παλινδρόμηση, δέντρα αποφάσεων, μηχανές διανυσμάτων υποστήριξης (*SVM*), ...

Γραμμική διαχωρισιμότητα

- **Όχι δεδομένη για κάθε πρόβλημα ταξινόμησης**
 - π.χ στο σχήμα δεξιά δεν μπορεί να βρεθεί υπερεπίπεδο (ευθεία γραμμή) που να χωρίζει ακριβώς το χώρο των δεδομένων σε δύο περιοχές, ανάλογα με την κλάση που ανήκουν



Θεώρημα του Cover

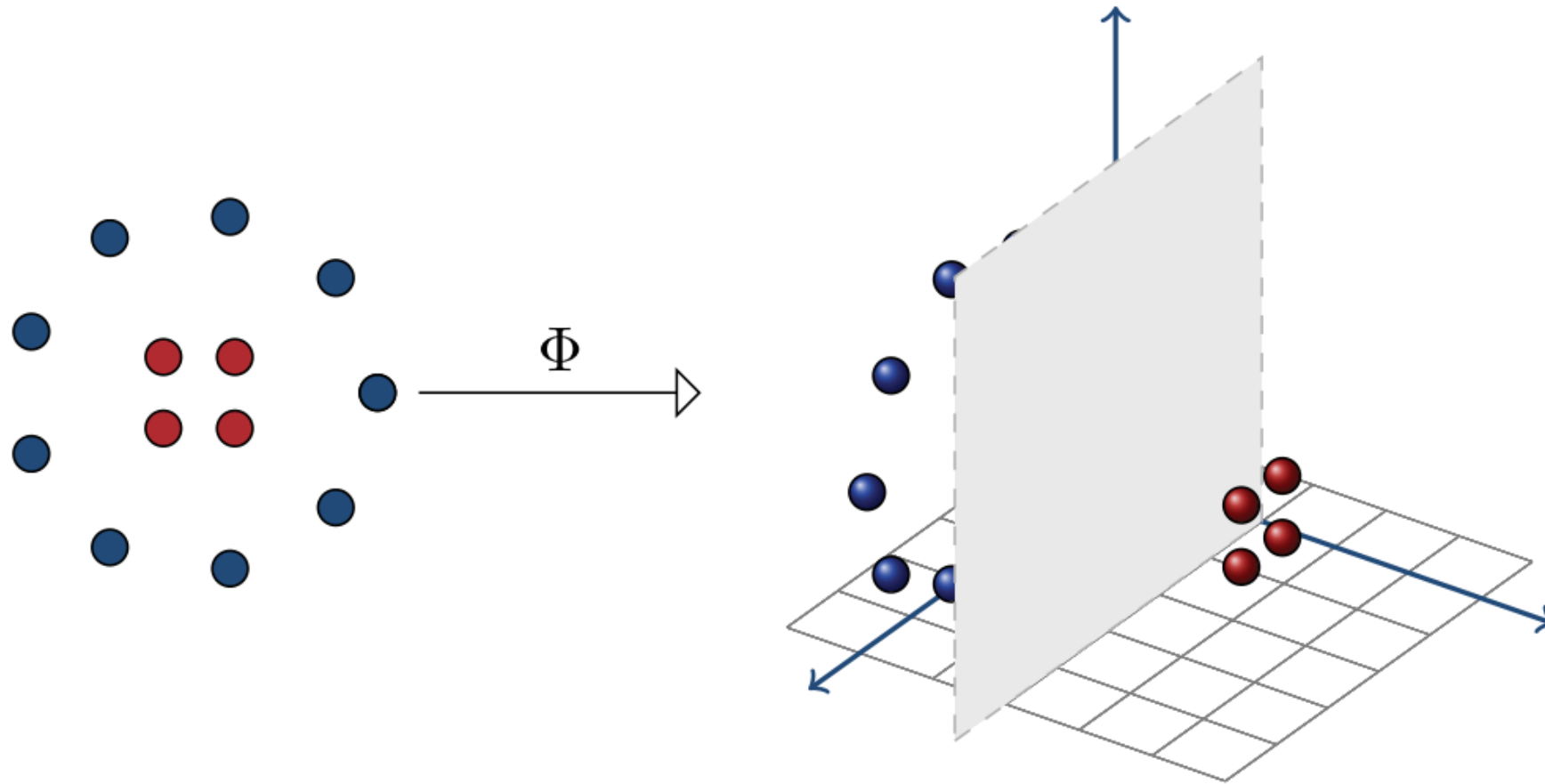
*A complex pattern-classification problem, cast in a **high-dimensional space nonlinearly**, is more likely to be **linearly separable** than in a low-dimensional space, provided that the space is not densely populated.*

Cover, T.M., *Geometrical and Statistical properties of systems of linear inequalities with applications in pattern recognition*, 1965

• Χαρακτηριστικά

1. Μη-γραμμική απεικόνιση
2. Χώρος μεγαλύτερων διαστάσεων

Θεώρημα του Cover - Παράδειγμα



- Στην πράξη, ο *χώρος της απεικόνισης* μπορεί να γίνει **πολύ μεγάλος**
 - Ζητήματα *υπολογιστικού κόστους, αποθήκευσης, ...*
- Η *ικανότητα γενίκευσης* ορισμένων ταξινομητών **δεν εξαρτάται** από τις *διαστάσεις* του χώρου εισόδου (χαρακτηριστικών) **αλλά** από **άλλες παραμέτρους**
 - πχ, για τις μηχανές διανυσμάτων υποστήριξης εξαρτάται από το **περιθώριο** (*margin*) μεταξύ των κλάσεων και το **πλήθος** των *δειγμάτων* εκπαίδευσης
- **Λύση**
 - Χρήση των **μεθόδων πυρήνα** (*kernel methods*) που βασίζονται στις **συναρτήσεις πυρήνα** (*kernel functions*) ή, απλούστερα, **πυρήνες** (*kernels*)
 - Διατηρείται η έννοια της **μη-γραμμικότητας**

Συναρτήσεις πυρήνα

- Η συνάρτηση $K: \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ καλείται **συνάρτηση πυρήνα** στον χώρο δεδομένων \mathcal{X} , όταν για οποιαδήποτε δύο σημεία $x, x' \in \mathcal{X}$, το $K(x, x')$ γίνεται **ίσο** με ένα **εσωτερικό γινόμενο** δύο διανυσμάτων $\Phi(x)$ και $\Phi(x')$
 - $\forall x, x' \in \mathcal{X}, K(x, x') = \langle \Phi(x), \Phi(x') \rangle$
 - $\Phi: \mathcal{X} \rightarrow \mathbb{H}$ **μη-γραμμική** απεικόνιση από τον χώρο των δεδομένων σε ένα **χώρο Hilbert** \mathbb{H} που καλείται **χώρος χαρακτηριστικών** (*feature space*)
- **Πλεονεκτήματα**
 1. **Αποτελεσματικότητα**
 - Πολύ πιο **αποδοτικός** ο υπολογισμός του K , απ' ότι του Φ ή ενός εσωτερικού γινομένου σημείων στο \mathbb{H}
 2. **Προσαρμοστικότητα**
 - **Δεν χρειάζεται** να οριστεί ή να υπολογιστεί **αναλυτικά** η Φ .
 - **Αυθαίρετη** επιλογή K , αρκεί να είναι **εγγυημένη** η ύπαρξη της Φ
 \Rightarrow η K πρέπει να ικανοποιεί το **Θεώρημα του Mercer**

Χώροι Hilbert

- David Hilbert (1862-1943)
- **Γενίκευση του Ευκλείδειου χώρου**
2 διαστάσεων σε χώρους (θεωρητικά)
απεριόριστων διαστάσεων
 - *Επέκταση των μεθόδων της γραμμικής άλγεβρας και του λογισμού*
- **Πλήρης (*complete*) μετρικός χώρος**
 - Έχει τέτοια όρια που **επιτρέπει** τη χρήση των μεθόδων του λογισμού
 - Είναι εφοδιασμένος με την πράξη του εσωτερικού γινομένου

Θεώρημα του Mercer

Έστω $\mathcal{X} \subset \mathbb{R}^N$ ένας συμπαγής χώρος δεδομένων N διαστάσεων και $K: \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ μια **συνεχής** και **συμμετρική** συνάρτηση. Τότε η K επιδέχεται ένα *ομοιόμορφο* και *συγκλίνων* ανάπτυγμα της μορφής

$$K(x, x') = \sum_{n=0}^{\infty} \alpha_n \varphi_n(x) \varphi_n(x'), \quad \alpha_n > 0$$

ανν ισχύει η παρακάτω **ανισότητα** για κάποια τετραγωνικά ολοκληρώσιμη συνάρτηση c ($c \in L_2(\mathcal{X})$)

$$\iint_{\mathcal{X} \times \mathcal{X}} c(x)c(x')K(x, x')dx dx' \geq 0$$

Θεώρημα Mercer - Συνέπειες

- Εφόσον K **συνεχής**, **συμμετρική** και ικανοποιεί το **ΘΜ**, τότε είναι **συνάρτηση πυρήνα**
 - Δε χρειάζεται να γνωρίζω την Φ
 - Μου αρκεί που γνωρίζω ότι *μπορεί* να γραφεί ως **εσωτερικό γινόμενο** μιας κάποιας **άγνωστης** Φ
 - Η ικανοποίηση της ανισότητας εξασφαλίζει την **κυρτότητα** (*convexity*) του προβλήματος της βελτιστοποίησης, όταν η K χρησιμοποιείται σε τεχνικές όπως τα SVM (σύγκλιση σε ολικά βέλτιστη λύση)
- Μια συνάρτηση K που είναι **συνεχής**, **συμμετρική** και **θετικά ορισμένη** (*positive definite symmetric – PDS*) ικανοποιεί το **ΘΜ**
 - *Πιο γενική* ιδιότητα
 - *Δεν απαιτεί* **καμία υπόθεση** για τον *χώρο* των *δεδομένων* \mathcal{X}

Συναρτήσεις πυρήνα

Ορισμός

Ένας πυρήνας $K: \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ καλείται **συνεχής, συμμετρικός και θετικά ορισμένος**, αν $\forall x_i, x_j \in \mathcal{X}$ ο πίνακας $\mathbf{K} = [K(x_i, x_j)]_{ij} \in \mathbb{R}^{m \times m}$ είναι **συμμετρικός και θετικά ημιορισμένος** (*symmetric positive semidefinite – SPSD*)

- Ο πίνακας \mathbf{K} είναι SPSD αν είναι **συμμετρικός** και **μια** από τις **δύο** ισοδύναμες συνθήκες ισχύουν
 1. Οι *ιδιοτιμές* του \mathbf{K} είναι *μη-αρνητικές*
 2. Για κάθε διάνυσμα $\mathbf{c} = (c_1, \dots, c_m)^\top \in \mathbb{R}^{m \times 1}$ ισχύει $\mathbf{c}^\top \mathbf{K} \mathbf{c} = \sum_{i=1}^n \sum_{j=1}^n c_i c_j K(x_i, x_j) \geq 0$
- Ο \mathbf{K} καλείται επίσης και **πίνακας Gram**

Παραδείγματα συναρτήσεων πυρήνα

- **Πολυωνυμικοί πυρήνες** (*Polynomial kernels*)

- Βαθμού $d \in \mathbb{N}$
- $\forall \mathbf{x}, \mathbf{x}' \in \mathbb{R}^N, K(\mathbf{x}, \mathbf{x}') = (\mathbf{x} \cdot \mathbf{x}' + c)^d$, όπου $c > 0$ σταθερά
- Απεικόνιση από χώρο N σε χώρο $\binom{N+d}{d}$

- **Γκαουσιανοί πυρήνες** (*Gaussian kernels*)

- Επίσης γνωστοί και ως **ακτινικές συναρτήσεις βάσης** (*radial basis functions – RBF*)
- $\forall \mathbf{x}, \mathbf{x}' \in \mathbb{R}^N, K(\mathbf{x}, \mathbf{x}') = e^{-\frac{\|\mathbf{x}-\mathbf{x}'\|^2}{2\sigma^2}}$, όπου $\sigma > 0$ σταθερά

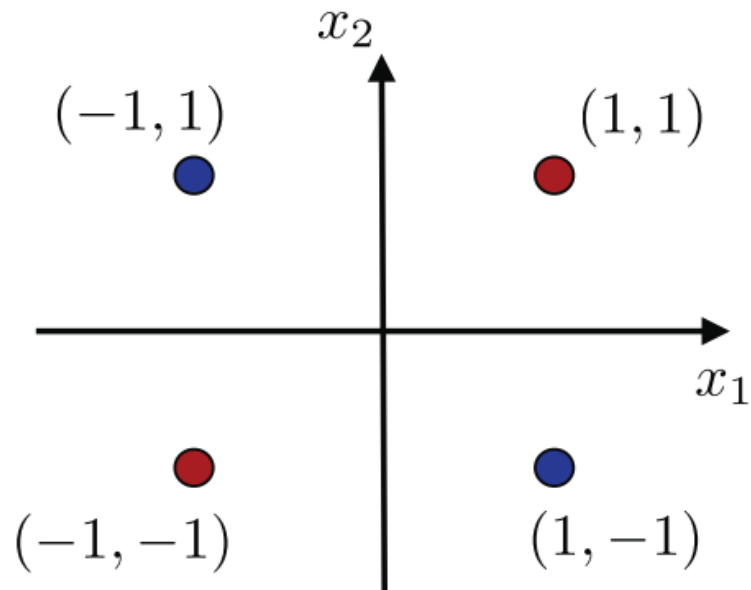
- **Σιγμοειδείς πυρήνες** (*sigmoid kernels*)

- $\forall \mathbf{x}, \mathbf{x}' \in \mathbb{R}^N, K(\mathbf{x}, \mathbf{x}') = \tanh(\alpha(\mathbf{x} \cdot \mathbf{x}') + b)$, όπου $\alpha, b \geq 0$ σταθερά
- Η εκπαίδευση των SVM με σιγμοειδείς πυρήνες ομοιάζει στη λειτουργία με τα πολυεπίπεδα Perceptron

Παράδειγμα πολυωνυμικού πυρήνα

- Έστω ότι έχουμε πρόβλημα ταξινόμησης σε χώρο $N = 2$ διαστάσεων *μη-γραμμικά διαχωρίσιμο*
- Χρησιμοποιούμε **πολυωνυμικό πυρήνα** 2^{ου} βαθμού ($d = 2$)
 - Διαστάσεις χώρου προβολής $\binom{2+2}{2} = 6$
 - $\forall \mathbf{x}, \mathbf{x}' \in \mathbb{R}^2$
 - $K(\mathbf{x}, \mathbf{x}') = (\mathbf{x} \cdot \mathbf{x}' + c)^2 = \left([x_1 \ x_2] \cdot \begin{bmatrix} x'_1 \\ x'_2 \end{bmatrix} + c \right)^2 = (x_1 x'_1 + x_2 x'_2 + c)^2$
 - $K(\mathbf{x}, \mathbf{x}') = [x_1^2 \ x_2^2 \ \sqrt{2}x_1x_2 \ \sqrt{2c}x_1 \ \sqrt{2c}x_2 \ c] \cdot [x'^2_1 \ x'^2_2 \ \sqrt{2}x'_1x'_2 \ \sqrt{2c}x'_1 \ \sqrt{2c}x'_2 \ c]^T$
 - $K(\mathbf{x}, \mathbf{x}') = \langle \Phi(\mathbf{x}), \Phi(\mathbf{x}') \rangle$
- $\Phi(\mathbf{x}) = [x_1^2 \ x_2^2 \ \sqrt{2}x_1x_2 \ \sqrt{2c}x_1 \ \sqrt{2c}x_2 \ c]$
 - $\Phi(\mathbf{x})$ **μη-γραμμική απεικόνιση** ενός δείγματος στο χώρο 6 διαστάσεων (απόδειξη ότι ο πολυωνυμικός πυρήνας είναι PD).
 - Περιλαμβάνει τα αρχικά χαρακτηριστικά (x_1, x_2) , τα γινόμενά τους καθώς και τη σταθερά c
 - Γενικότερα, για πολυωνυμικό πυρήνα d βαθμού έχουμε *απλούς όρους μέχρι d βαθμού*

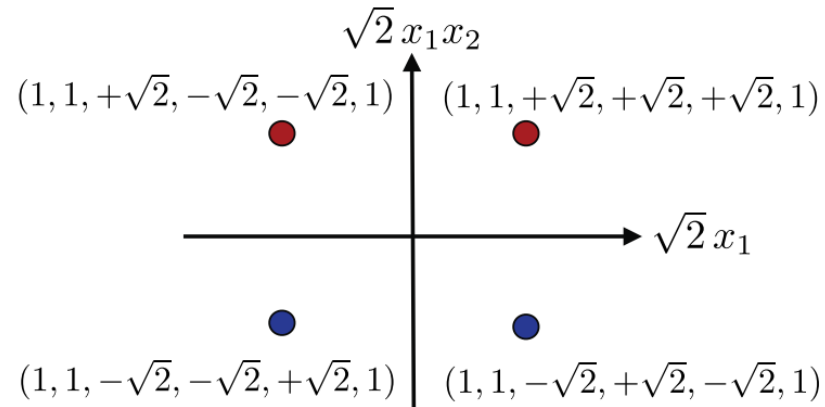
Παράδειγμα – Πρόβλημα XOR



- Μη-γραμμικά διαχωρίσιμο πρόβλημα 2 διαστάσεων
 - Διπολική μορφή
- Χρήση πολυωνυμικού πυρήνα 2^{ου} βαθμού
 - Όπως προηγουμένως

Πρόβλημα XOR (συνέχεια)

- $\Phi(x) = [x_1^2 \ x_2^2 \ \sqrt{2}x_1x_2 \ \sqrt{2}cx_1 \ \sqrt{2}cx_2 \ c]$
 - Προβολή σε χώρο 6 διαστάσεων
 - Γραμμικά διαχωρισίμο ως προς υπερεπίπεδο $x_1x_2 = 0$
 - Στο αριστερά σχεδιάγραμμα απεικονίζεται προβολή ως προς την 3^η και την 4^η διάσταση



Reproducing kernel Hilbert Space (RKHS)

- **Θεώρημα RKHS**
 - Έστω PDS πυρήνας $K: \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$. Τότε **υπάρχει χώρος Hilbert \mathbb{H} και απεικόνιση Φ έτσι ώστε να ισχύουν:**
 1. $\forall x, x' \in \mathcal{X}, K(x, x') = \langle \Phi(x), \Phi(x') \rangle$
 2. $\forall h \in \mathbb{H}, \forall x \in \mathcal{X}, h(x) = \langle h, K(x, \cdot) \rangle$
 - Ο χώρος \mathbb{H} καλείται reproducing kernel Hilbert space σχετιζόμενος με τον πυρήνα K
- Το θεώρημα RKHS έχει ως **συνέπεια** ότι κάθε PDS πυρήνας **ορίζει** ένα **χώρο χαρακτηριστικών**
 - Άρα, σε αυτή την περίπτωση, το *ζητούμενο* της μάθησης μεταφράζεται από την εύρεση «καλών» χαρακτηριστικών στην εύρεση «καλών» πυρήνων PDS

Κανονικοποιημένοι πυρήνες

- Σε κάθε πυρήνα K αντιστοιχεί **κανονικοποιημένος πυρήνας** K' , όπου

$$K'(x, x') = \begin{cases} 0, & \text{if } K(x, x) = 0 \text{ or } K(x', x') = 0 \\ \frac{K(x, x')}{\sqrt{K(x, x)K(x', x')}} & \text{otherwise} \end{cases}$$

- Εξ' ορισμού, αν $K'(x, x) \neq 0$ τότε $K'(x, x) = 1, \forall x \in \mathcal{X}$
- Αν K είναι πυρήνας PDS, τότε και K' PDS
- Οι **κανονικοποιημένοι** πυρήνες μπορούν να ερμηνευτούν ως **μετρικές ομοιότητας**
 - Η μορφή τους εκφράζει το **συνημίτονο** της γωνίας μεταξύ $\Phi(x)$ και $\Phi(x')$
- **Παράδειγμα** κανονικοποιημένου πυρήνα αποτελεί ο **γκαουσιανός**
 - $K_1(x, x') = e^{-\frac{\|x-x'\|^2}{2\sigma^2}}$ κανονικοποιημένος πυρήνας του $K_2(x, x') = e^{-\frac{x \cdot x'}{2\sigma^2}}$

Εμπειρική Απεικόνιση πυρήνα

- Σε ορισμένες περιπτώσεις, επιθυμούμε να **ορίσουμε** εμείς μια **απεικόνιση** και επιπρόσθετα θέλουμε αυτή να **αντιστοιχεί** σε έναν **PDS πυρήνα**
 - λ.χ. θέλουμε να αποτυπώσουμε τη γνώση που έχουμε σε μια συγκεκριμένη περιοχή (*domain knowledge*)
- Έστω ότι έχουμε m δείγματα δεδομένων σε χώρο \mathcal{X} ($x_1 \cdots x_m \in \mathcal{X}$)
 - Η **εμπειρική απεικόνιση πυρήνα** (*empirical kernel map*) $\Phi: \mathcal{X} \rightarrow \mathbb{R}^m$ που σχετίζεται με PDS πυρήνα K ορίζεται ως το διάνυσμα
 - $\forall x \in \mathcal{X}, \Phi(x) = [K(x, x_1) \cdots K(x, x_m)]$
 - Είναι το διάνυσμα των **μετρικών ομοιότητας** K κάθε σημείου του χώρου με **όλα** τα υπόλοιπα
 - Αποδεικνύεται ότι αν \mathbf{K} είναι ο πίνακας πυρήνα του K , τότε \mathbf{K}^2 είναι ο πίνακας πυρήνα της απεικόνισης Φ

Ιδιότητες κλειστότητας (closure)

- Οι πυρήνες PDS είναι **κλειστοί** (*closed*) ως προς το **άθροισμα**, το **γινόμενο**, το **γινόμενο Hadamard**, τη **σύγκλιση κατά σημείο** (όριο) και τη **σύνθεση δυναμοσειρών** της μορφής $\sum_{n=0}^{\infty} a_n x^n$, $a_n \geq 0, \forall n \in \mathbb{N}$
- Χρήσιμη ιδιότητα γιατί *αν γνωρίζουμε* PDS πυρήνες για *υπο-προβλήματα*, μπορούμε να τους **συνδυάσουμε**, γνωρίζοντας ότι ο **συνδυασμός** τους είναι *επίσης* PDS πυρήνας

Αλγόριθμοι
ταξινόμησης
βασισμένοι σε
πυρήνες

SVM με πυρήνες PDS

Πρόβλημα βελτιστοποίησης SVM

- $\max_{\lambda} \left[\sum_{i=1}^m \lambda_i - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \lambda_i \lambda_j d_i d_j x_i^\top x_j \right]$ ως προς περιορισμούς
- $\sum_{i=1}^m \lambda_i d_i = 0$ και $\forall i \lambda_i \in [0, C]$
- όπου m πλήθος δειγμάτων, d_i ετικέτα i -οστού δείγματος x_i , λ_i πολλαπλασιαστές Lagrange και C βάρος κόστους λανθασμένων λύσεων
- Αντικατάσταση **εσωτερικού γινομένου** $x_i^\top x_j$ από πυρήνα $K(x_i, x_j)$
 - Εξάλλου και ο πυρήνας K ορίζει πράξη εσωτερικού γινομένου σε χώρο απεικόνισης Φ
 - $\max_{\lambda} \left[\sum_{i=1}^m \lambda_i - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \lambda_i \lambda_j d_i d_j K(x_i, x_j) \right]$
- SVM με πυρήνες PDS (υπό μορφή πινάκων)
 - $\max_{\lambda} [\mathbf{1}^\top \lambda - (\lambda \circ \mathbf{d})^\top \mathbf{K}(\lambda \circ \mathbf{d})]$ ως προς περιορισμούς
 - $\lambda^\top \mathbf{d} = \mathbf{0}$ και $\mathbf{0} \leq \lambda \leq \mathbf{C}$
 - \circ πολλαπλασιασμός Hadamard

Θεώρημα Representer

Έστω PDS πυρήνας $K: \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ και \mathbb{H} ο RKHS χώρος του. Τότε, για κάθε μη-φθίνουσα συνάρτηση $G: \mathbb{R} \rightarrow \mathbb{R}$ και για κάθε συνάρτηση σφάλματος $L: \mathbb{R}^m \rightarrow \mathbb{R} \cup \{+\infty\}$, το πρόβλημα βελτιστοποίησης

$$\operatorname{argmin}_{h \in \mathbb{H}} F(h) = \operatorname{argmin}_{h \in \mathbb{H}} G(\|h\|_{\mathbb{H}}) + L(h(x_1), \dots, h(x_m))$$

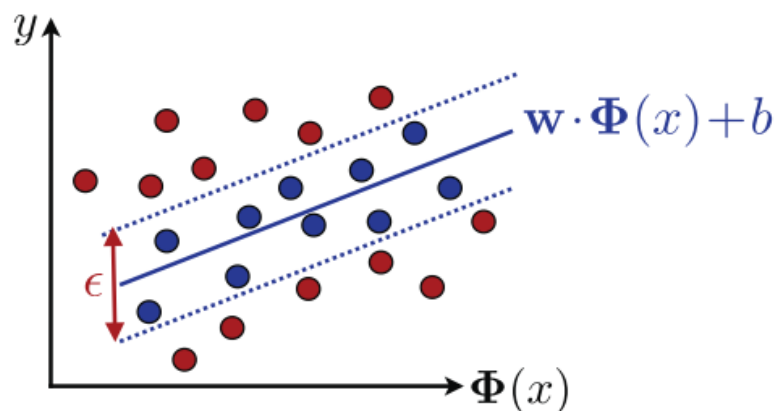
επιδέχεται λύση της μορφής $h^* = \sum_{i=1}^m \lambda_i K(x_i, \cdot)$. Επιπρόσθετα, αν G αύξουσα, τότε κάθε λύση έχει αυτή τη μορφή.

- Συνέπεια του θεωρήματος Representer είναι ότι ο **χώρος λύσεων** των **προβλημάτων βελτιστοποίησης** είναι ο **γραμμικός** συνδυασμός των συναρτήσεων $K(x_i, \cdot)$
 - Άρα και των SVM
 - x_i δείγμα δεδομένων

- **Πολυπλοκότητα Rademacher** για τον χώρο υποθέσεων των πυρήνων
 - Έστω PDS πυρήνας $K: \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ και $\Phi: \mathcal{X} \rightarrow \mathbb{H}$ η σχετιζόμενη απεικόνιση στο χώρο των χαρακτηριστικών.
 - Έστω $S \subseteq \{x: K(x, x) \leq r^2\}$ και έστω $\mathcal{H} = \{x \mapsto \langle \mathbf{w}, \Phi(x) \rangle: \|\mathbf{w}\|_{\mathbb{H}} \leq \Lambda\}$ για $\Lambda \geq 0$.
 - Τότε $\widehat{\mathfrak{R}}_S(\mathcal{H}) \leq \frac{\Lambda \sqrt{\text{TR}[K]}}{m} \leq \sqrt{\frac{r^2 \Lambda^2}{m}}$
- Το **ίχνος** (*trace*) του πίνακα πυρήνα K καθορίζει την **πολυπλοκότητα** του χώρου των υποθέσεων

Αλγόριθμοι
παλινδρόμησης
βασισμένοι σε
πυρήνες

Παλινδρόμηση Διανυσμάτων Υποστήριξης



- **Support vector regression (SVR)**
- Σε προβλήματα παλινδρόμησης επιθυμούμε να βρούμε την *υποκείμενη συνάρτηση* που παράγει τα δεδομένα
- Εδώ επιθυμούμε να βρούμε *περιοχή ϵ γύρω* από την **υποκείμενη** συνάρτηση
- Τα σημεία του χώρου χωρίζονται σε **2 περιοχές**
 - Σε αυτά που είναι **εντός** ϵ από την υπολογιζόμενη συνάρτηση (μπλε)
 - Σε αυτά που είναι **μακρύτερα** από ϵ από την υπολογιζόμενη συνάρτηση (κόκκινα)
 - Σε αυτά εφαρμόζεται «**ποινή**», ανάλογα με την απόστασή τους από την υπολογιζόμενη συνάρτηση

Παλινδρόμηση Διανυσμάτων Υποστήριξης

- Για χώρο υποθέσεων γραμμικών συναρτήσεων $\mathcal{H} = \{x \mapsto \mathbf{w} \cdot \Phi(x) + b : \mathbf{w} \in \mathbb{R}^N, b \in \mathbb{R}\}$, όπου Φ απεικόνιση χαρακτηριστικών που αντιστοιχεί σε PDS πυρήνα K , το πρόβλημα βελτιστοποίησης του SVR παίρνει τη μορφή
 - $\min_{\mathbf{w}, b} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^m |y_i - (\mathbf{w} \cdot \Phi(x_i) + b)|_\epsilon$
 - Το $|\cdot|_\epsilon$ δηλώνει ϵ -«αναισθησία» στην απώλεια (ϵ -insensitive loss)
 - $\forall y, y' \in \mathcal{Y}, |y' - y|_\epsilon = \max(0, |y' - y| - \epsilon)$
- Με τη χρήση μεταβλητών χαλαρότητας $\xi_i \geq 0$ και $\xi'_i \geq 0$, το πρόβλημα βελτιστοποίησης γίνεται
 - $\min_{\mathbf{w}, b, \xi, \xi'} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^m (\xi_i + \xi'_i)$, υπό περιορισμούς
 - $(\mathbf{w} \cdot \Phi(x_i) + b) - y_i \leq \epsilon + \xi_i$
 - $y_i - (\mathbf{w} \cdot \Phi(x_i) + b) \leq \epsilon + \xi'_i$

Παλινδρόμηση Διανυσμάτων Υποστήριξης

- Το πρόβλημα που περιγράφηκε στην προηγούμενη διαφάνεια είναι **κυρτό τετραγωνικό** (*convex quadratic*) με **αφινικούς περιορισμούς** (*affine constraints*) και με την χρήση των *πολλαπλασιαστών Lagrange* και την *ικανοποίηση των συνθηκών Kuhn-Karush-Tucker* (KKT), όπως ισχύει και για τα SVM, οδηγεί στο ακόλουθο ισοδύναμο δυικό πρόβλημα (χρήση πίνακα πυρήνα **K**)
 - $\max_{\lambda, \lambda'} -\epsilon(\lambda + \lambda')^\top \mathbf{1} + (\lambda' - \lambda)^\top \mathbf{y} - \frac{1}{2} (\lambda' - \lambda)^\top \mathbf{K} (\lambda' - \lambda)$, υπό περιορισμούς
 - $(\lambda + \lambda')^\top \mathbf{1} = 0, \mathbf{0} \leq \lambda \leq \mathbf{c}, \mathbf{0} \leq \lambda' \leq \mathbf{c}$
- Κάθε PDS πυρήνας **K** μπορεί να χρησιμοποιηθεί σε SVR
 - Άρα ο αλγόριθμος επεκτείνεται και σε **μη-γραμμικές παλινδρομήσεις**

Παλινδρόμηση Ridge με χρήση πυρήνων

- **Kernel Ridge Regression**
 - Τεχνική ομαλοποίησης (*regularization*), αντίστοιχη της L2 ομαλοποίησης στα προβλήματα ταξινόμησης
 - $\min_w F(w) = \sum_{i=1}^m (w \cdot \Phi(x_i) - y_i)^2 + \lambda \|w\|^2$
 - λ παράμετρος που «ελέγχει» τη συνεισφορά του όρου ομαλοποίησης στη συνάρτηση κόστους (εδώ το μέσο τετραγωνικό σφάλμα)
 - Υπολογίζουμε την κλήση της F ως προς w και την εξισώνουμε με το μηδέν για να βρούμε τις βέλτιστες τιμές των όρων της εξίσωσης

Παλινδρόμηση Ridge με χρήση πυρήνων

- **Εναλλακτική μορφή**
 - $\min_w \sum_{i=1}^m (w \cdot \Phi(x_i) - y_i)^2$, υπό περιορισμό $\|w\|^2 \leq \Lambda^2$
- Χρήση **συνθηκών KKT** και **θεωρήματος Representer** οδηγούν σε λύσεις της μορφής $\forall x \in \mathcal{X}, h(x) = w \cdot \Phi(x) = \sum_{i=1}^m \lambda_i K(x_i, x)$

Βιβλιογραφία

1. M. Mohri, A. Rostamizadeh, A. Talwalker – Foundations of Machine Learning
 - Κεφάλαιο 6
 - Ενότητα 11.3.2
 - Ενότητα 11.3.3
2. S. Shalev-Shwartz, S. Ben-David – Understanding machine learning: From theory to algorithms
 - Κεφάλαιο 16